To appear in the *International Symposium of Robotics Research*, Munich, Germany, October 1995.

Obstacle Detection for Unmanned Ground Vehicles: **A Progress Report***

Larry Matthies, Alonzo Kelly, and Todd Litwin
Jet Propulsion Laboratory - California Institute of Technology
4800 Oak Grove Drive
Pasadena, California 91109

May 11, 1995

EXTENDED ABSTRACT

* The work described in this paper was sponsored by the Advanced Research Projects Agency.

1 Introduction

Unmanned ground vehicles (UGV's) are being developed for a variety of applications in the military, in hazardous waste remediation, and in planetary exploration. Such applications often involve limitations on communications that require UGV's to navigate autonomou sl y for extended distances and extended periods of time. Under these circumstances, UGV'S must be equipped with sensors for detecting obstacles in their path. This paper provides a progress report on work being done at the Jet Propulsion Laboratory (JPL) on obstacle detection sensors for the "Demo II" UGV Program, which is sponsored by the Advanced Research Projects Agency (ARPA) and the Office of the Secretary of Defense (OS D).

The goal of the Demo II Program is to develop technology enabling UGV'S to perform autonomous scouting missions. In its full generality, this application will require operating during the day or night, in clear or obscured weather conditions, and over terrain that will include a variety of natural and man-made obstacles. Obstacle detection sensors for the full problem must be able to perceive the terrain geometry, perceive the material type of any ground cover (ie. terrain type), and do so at night and through haze.

We have been addressing the problem of perceiving terrain geometry (by day and by night) and we are beginning to address the problem of perceiving terrain type (by day). For sensing geometry, we are using stereo vision, because its properties of being non-emissive, non-scanning, and non-mechanical make it attractive for military vehicles that require a low signature sensor and well-registered range data while jostling over rough terrain. In collaboration with other labs, we have recently begun experimenting with stereo vision on thermal imagery to address night operations. For sensing terrain type, we have done preliminary investigations of discriminating soil, vegetation, and water using visible, near infrared, and polarization imagery.

Section 2 reviews the current status of the real-time stereo vision system we have developed for UGV applications.> The system currently produces range imagery from a 256 x 45 pixel window of attention in about 0.6 seconds/frame, using a Datacube MV-200 image processing board and a 68040 CPU board as the sole computing engines. This vision system is installed on a roboticized HMMWV[1] that serves as a testbed UGV. Section 3 describes a number of enhancements currently in progress on this system, including recent tests with thermal imagery, simple algorithms for real-time obstacle detection, methods to support focus of attention, and approaches to terrain classification. Section 4 reviews how the vision system has evolved through three major demonstrations over the last five years, and shows results from an autonomous navigation trial conducted with our HMMWV on a dirt road near the laboratory. These demonstrations have driven HMMWV'S at speeds in the neighborhood of 5 to 10 kph over gentle, but not barren cross-country terrain.

This work has been the first to show that stereo vision can provide range data of sufficient quality, at sufficient speeds, and with sufficiently small computing resources to be practical for UGV navigation. Future work will attempt to increase the quality of the range data, miniaturize the computing system, and integrate terrain classification with range imaging,

[1] High Mobilit y Multipurpose Wheeled Vehicle - the modern military jeep.

2 The JPL Stereo Vision System

Previous versions of JPL's real-time stereo vision system have been described in [1, 2]. Here, we will outline the current version of the algorithm, then discuss how and why it has changed. Current steps in the algorithm are:

1. Digitize fields of the stereo image pairs.
2. Rectify the fields.
3. Compute image pyramids by a difference-of-Gaussian image pyramid transformation.
4. Measure image similarity by computing the sum-squared-difference (S SD) for 7 x 7 windows over a fixed disparity search range.
5. Estimate disparity by finding the SSD minimum independently for each pixel.
6. Filter out bad matches using the left-right-line-of-sight (LRLOS) consistency check [3, 4].
7. Estimate sub-pixel disparity by fitting parabolas to the three SSD values surrounding the SSD minimum and taking the disparity estimate to be the minimum of the parabola,
8. Smooth the disparity map with a 3 x 3 low-pass filter to reduce noise and artifacts from the sub-pixel estimation process.
9. Filter out small regions (likely bad matches) by applying a blob filter that uses a threshold on the disparity gradient as the connectivity criterion.
10. Triangulate to produce the X-Y-Z coordinates at each pixel and transform to the vehicle coordinate frame.
11. Detect "positive" obstacles[2] by thresholding the output of a simple slope operator applied to the range image.

Since the first version, this algorithm has evolved as follows. The original version digitized full frames instead of fields (step 1), because it was first used on a Mars rover prototype vehicle that stopped to acquire imagery. At the time, the remainder of the algorithm consisted of steps 3, 4, 5, 7, 8, and 10, plus a Bayesian posterior probability measure that was applied in place of step 6 above to filter out bad matches. Matching was done at only at the 64 x 60-pixel level of resolution in the image pyramid. Changes to the system and the rationale behind them are described below.

Digitizing fields. Since the vehicle now drives continuously, fields are digitized instead of frames to avoid temporal misregistration from the field-interlaced cameras used on the vehicle.

[2] Positive obstacles are those that extend upward from the nominal ground plane, like rocks, bushes, and fence posts. "Negative" obstacles extend downward, like potholes, man-made ditches, and natural ravines.

Rectification. With the 64 x 60 resolution and 30 degree field of view (FOV) of the original system, it was possible to align the cameras well enough for stereo matching by mounting the cameras on mechanical y adjustable brackets, although the alignment procedure was painstaking. Since then, we have moved to matching at the 256 x 240 level of resolution and have used FOV'S as wide as 85 degrees. These changes produce sufficiently high angular resolution and radial distortion to make mechanical alignment impractical. Rectification currently works well enough that the residual vertical disparity at 256x 240 resolution is less than one pixel. Since the adjustable camera mounts had potential to go out of alignment, this approach is more robust, as well as much easier to use.

LRLOS consistency check. This procedure ensures that disparities obtained by choosing best matches along left camera lines of sight agree with those obtained by choosing matches along right camera lines of sight. We have found that this procedure runs faster than the posterior probability measure and works better, especially at occluding boundaries.

Blob **filter.** Small islands of bad matches occasionally survive the LRLOS test. These can be contiguous with larger regions of good matches or completely disjoint from good regions. A simple blob coloring algorithm [5] can be used to reject small disjoint regions. Furthermore, it is possible to eliminate small regions of bad matches that are cent iguous to larger, good regions by making the connectivity criterion in the blob coloring algorithm be a threshold on the disparity gradient.

Positive obstacle detection. As discussed elsewhere [6], obstacles can be difficult to define and expensive to compute. To enable real-time operation, we initially implemented an algorithm that detected only positive obstacles using a simple slope estimation algorithm applied to columns of the range image [2]. This algorithm assumes that obstacles are vertical step displacements of minimum height H on an otherwise flat ground plane. On each scanline, we project a line of sight out to intersect the ground plane, assume an obstacle of height H at that point, and determine on what scanline the top of the obstacle would appear in the image. A table of vertical scanline offsets is determined in this way for every pixel in the image. To detect obstacles, for every pixel p_1 in the range image, we use it and its higher partner p_2 in the same column to compute the change in height at that point in the image; if the change in height exceeds a threshold, the entire interval of the range image from p_1 to p_2 is declared to contain an obstacle. Although this algorithm is very simple, it is reasonably robust and very fast.

On our HMMWV, these algorithms are implemented in a computing system comprised of a Datacube MV-200 image processing board, a 68040 CPU board that controls the MV-200 over the VME bus, and a Sparcbook II notebook computer that communicates with the 68040 over ethernet, Figure 1 show how the major units of the algorithm map onto each processor, together with approximate run-times for processing a 45 scanline window at the 256x 240 level of resolution.

Figure 2 shows range data produced by the system. There area few gaps in the disparity map in the foreground, but no gross errors are evident. The rock in the middle of the road was 43 cm high and was placed 10 m from the cameras; it is perceived as approximately **42** cm high and 9 m from the cameras. The road is seen to be quite smooth up to the rock; bumps behind the rock may correspond to bushes on the side of the road. The elevation map shows a relatively smooth area

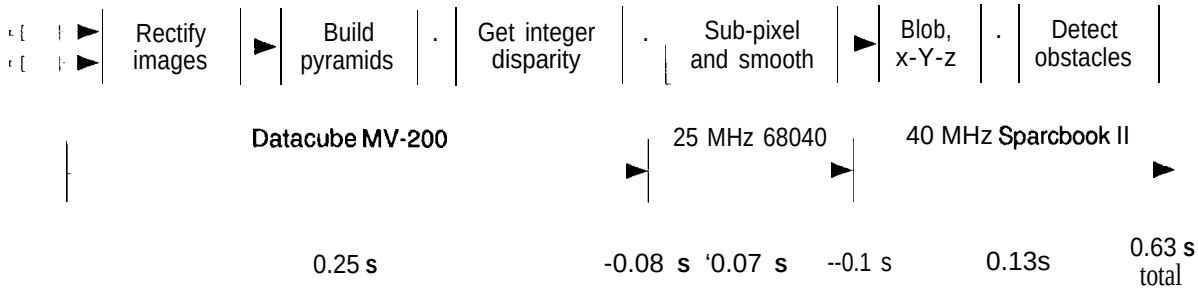


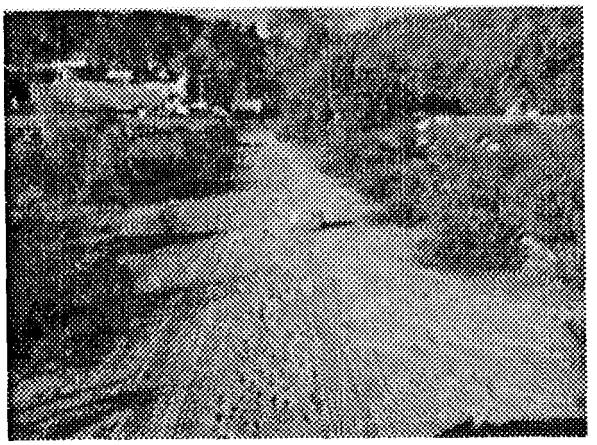
Figure 1: Major functional units of the stereo algorithm, their mapping onto hardware, and approximate run-times for each processor and 1/0 between the processors. Run-times are for a 256 x 45 window processed at the 256 x 240 level of resolution, with a disparity search range of 32 pixels.

corresponding to the roadway, high areas corresponding to the bushes beside the road, and a small bump corresponding to the rock on the road. Overall, the results are quite good.

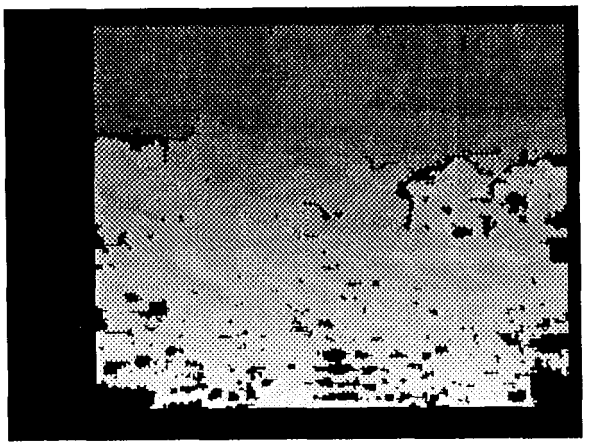
3 Ongoing Algorithm Extensions

A number of extensions to the basic stereo system are in progress. These have the goals of enabling night driving by using forward looking infrared (FLIR) cameras, detecting negative obstacles, increasing speed by using windows of attention and dual FOV camera systems, providing software exposure control, increasing the resolution by moving to full frame cameras, and augmenting obstacle detection by doing terrain classification as well as range imaging. These are listed in order from the most simple to the most complex. Initial implementations and results have been obtained for several of these functions; full frame imaging and terrain classification are not yet integrated into the real-time system. The current status for each issue is summarized below.

FLIR imagery. UGV'S must be able to perform missions at night. Stereo vision with thermal imagery obtained from FLIR cameras is a possible way to detect obstacles at night. The Demo 11 Program has acquired several FLIR cameras operating in the 3 to 5 μm wavelength range with 256 x 256 detectors. In collaboration with SRI International, a pair of these cameras were mounted on the JPL HMMWV and used for a preliminary evaluation of real-time range imaging. The cameras were mounted with a 30 cm baseline and had a field of view of approximately 30 degrees; this compares with 35 cm and 57 degrees for the visible imagery shown in figure 2. Figure 3 shows results obtained with FLIR imagery of the same scene as figure 2, processed at 128 x 120 resolution. Note that this data was processed with the *identical* algorithms used for the visible imagery. Qualitatively, the results compare quite well to the visible imagery. Results at 256 x 240 resolution are much sparser, for reasons yet to be determined. These initial results are encouraging, An important next step will be to characterize the quality of the range data quantitatively, in terms of the thermal sensitivity y of the cameras and temperature variations of the scene, at various times of the day and night.

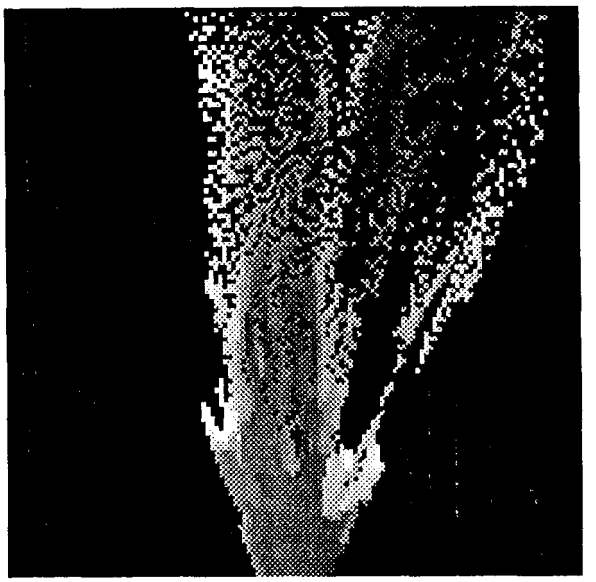


(a) Intensity image



(b) Disparity map

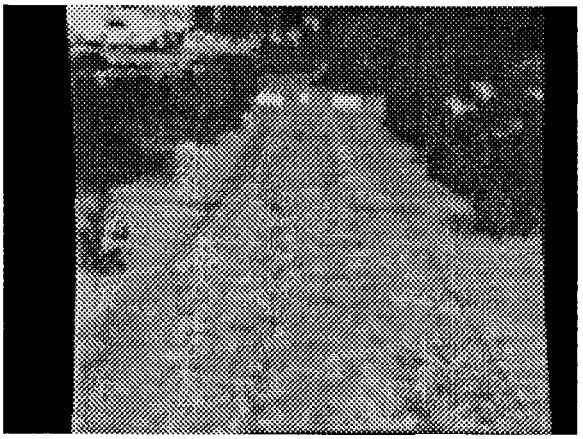
(c) Elevation profile through rock



(d) Elevation map

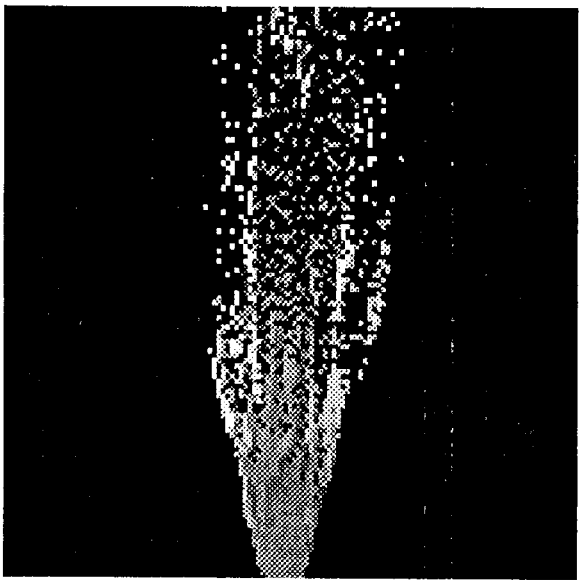
Figure 2: Range data for a dirt road in the Arroyo Seco, next to JPL. The disparity map was generated at 256 x 240 resolution; bright is close, dark is far. The elevation profile in (c) shows range pixels for a column through the rock in the middle of the road; the scale is 4 to 35 m horizontally and roughly -1 to +1.5 m vertically. In the elevation map, bright is high, dark is low; the scale is 4 to 35 m downrange and roughly -15 to +15 m crossrange.

(a) Left image



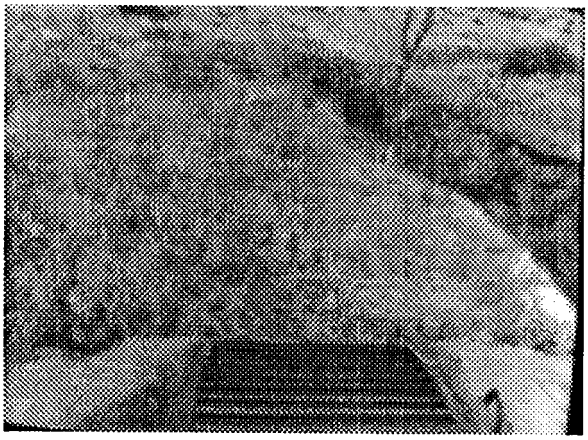
(b) Right image

(c) Disparity map

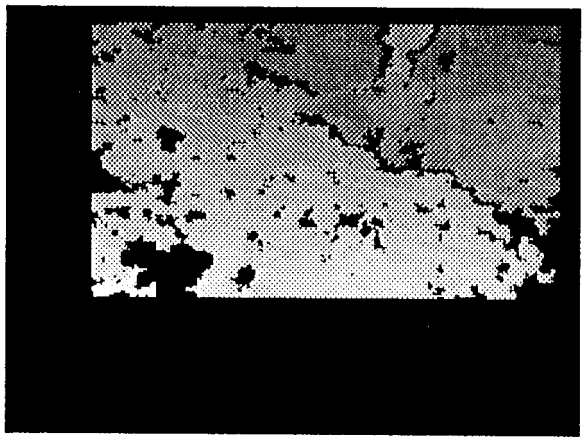


(d) Elevation map

Figure 3: Range data obtained with stereo FLIR imagery of the same road as figure 2. The disparity map was generated at 128 x 120 resolution. The scale of the elevation map matches figure 2.

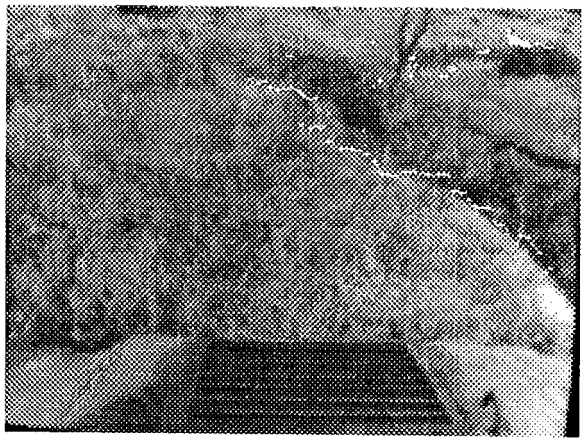


(a) Scene with tall grass and ravine

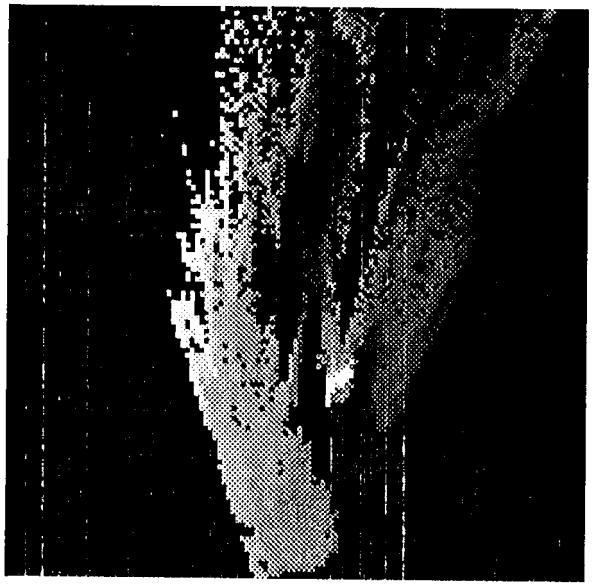


(b) Disparity map

(c) Positive obstacles (white overlay)



(d) Negative obstacles (white overlay)



(e) Elevation map

(f) Elevation map with obstacles (in white)

Figure 4: Obstacle detection results. See text for discussion.

to filter these quantities overtime and use them to drive lens aperture and/or image exposure time.

Experiments with software exposure control will be conducted in the spring and summer of 1995.

Full frame, non-interlaced range imaging. One of the least satisfactory characteristics of the stereo system is that it must discard half the available vertical resolution, due to the interlaced scanning of the cameras. However, cameras with full frame, non-interlaced scanning are now available (eg. the Pulnix TM-9700). The stereo software can be configured easily to process this imagery; however, three potential problems lurk:

- there will be more demanding requirements for image alignment because of the higher angular resolution of the images;
- there may be more matching ambiguity because of wider disparity ranges;
- there may be slower run-times because of the higher resolution.

We expect to address the first issue, if necessary, through improved calibration. One approach to the remaining issues is to pre-warp the imagery to produce zero disparity for the nominal ground plane; provided that no part of the scene deviates substantially from that plane, this technique reduces ambiguity and run-time by reducing the necessary disparity search range. Unfortunately, this is a fragile assumption, which will often be violated in cross-country navigation. However,

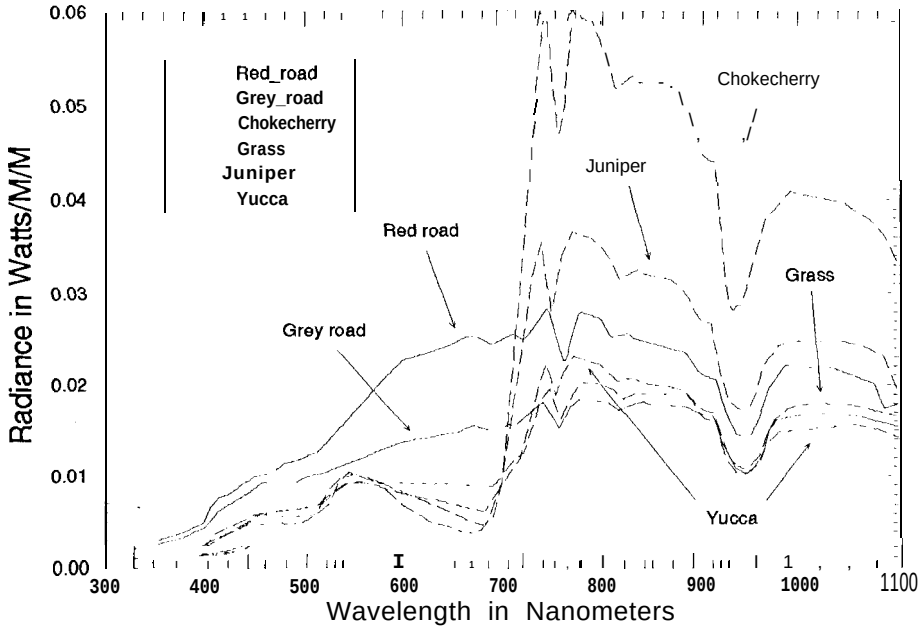


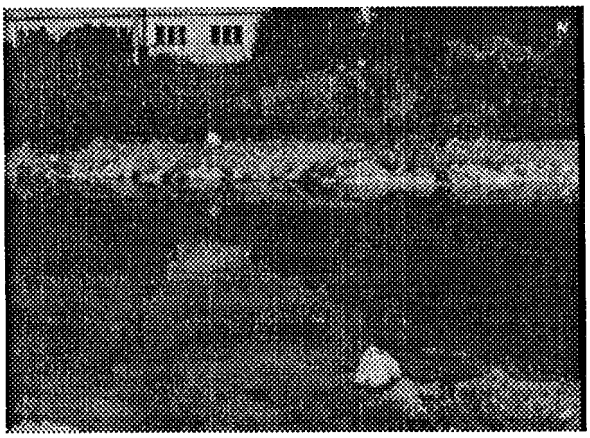
Figure 5: Spectral radiance for six terrain types, showing the generally brighter response of vegetation in the near infrared.

Figure *6* makes this point further by showing ground level imagery taken with a CCD camera and blue, green, red, and near infrared filters. The scene contains a pond in the foreground and various soils, bushes, trees, and buildings in the background. Although there are differences between the visible bands, they are much less dramatic than the contrast reversal shown by the soil and vegetation in the near infrared band. The 2-D histograms in figures 6e and 6f show that the visible bands tend to vary along a single brightness axis; however, with the near infrared band there is a separation: vegetation pixels tend to lie along the more vertical axis, while non-vegetation pixels tend to lie along the diagonal axis. We have used red, near infrared, and polarization imagery of this scene to obtain good discrimination between vegetation, water, and "other" pixels, using very simple, pixel-wise ratios of the bands. The results are dependent on the scene, however, and even more dependent on the season; in winter, with dead vegetation, the near infrared cue is unavailable. With other summertime imagery that has been segmented manually to estimate ground truth, we have seen improvements of up to 15% in correct classification rates using red and near infrared imagery, as compared to using red and green imagery. We plan to build simple, real-time implementations of both approaches in order to determine quickly whether a useful difference remains when cameras are panned across a wide range of scenery.

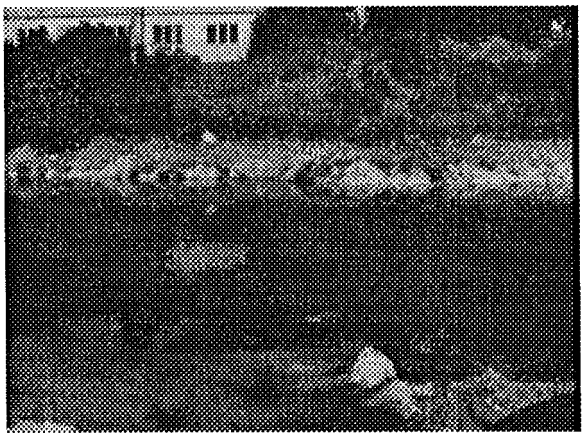
4 System Integration and Demonstrations

The stereo vision system described in section 2 has been integrated into robotic vehicles and used in the following configurations and demonstrations:

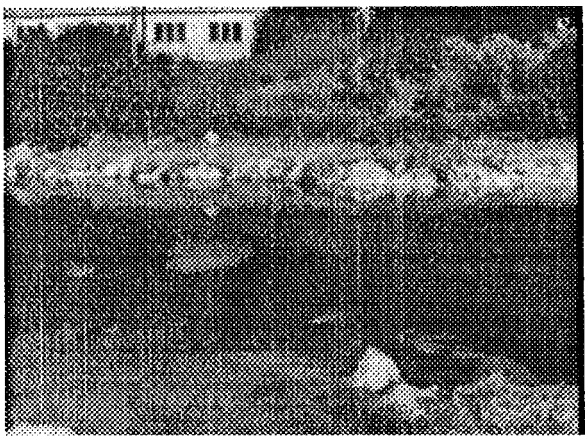
- On the Mars rover tested vehicle "Robby", running at 64 x 60 resolution in several seconds per frame, for autonomous navigation over 100 m of sandy terrain interspersed with bushes



(a) Blue band (450 nm)

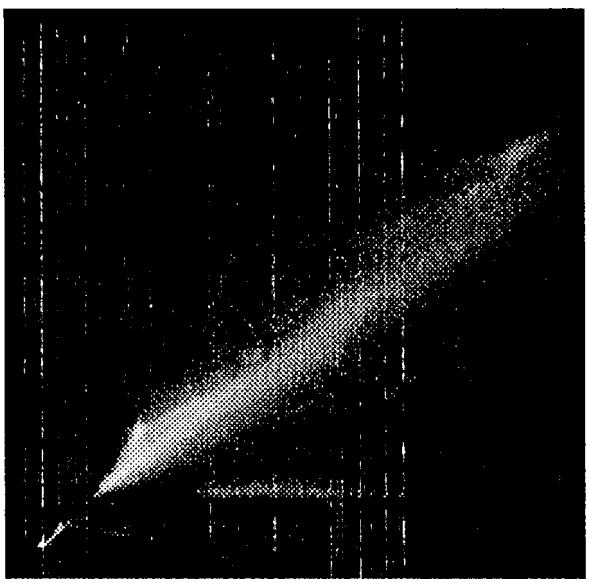


(b) Green band (550 nm)

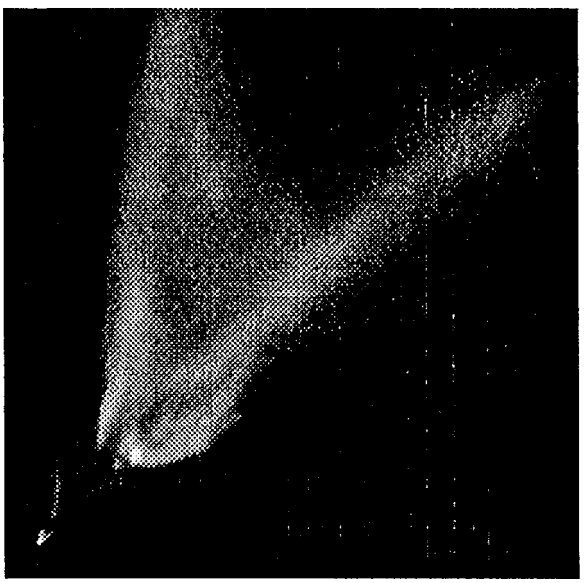


(c) Red band (650 nm)

(d) Near infrared band (800 nm)



(e) 650 vs. 550 histogram



(f) 650 vs. 800 histogram

Figure 6: Multispectral images

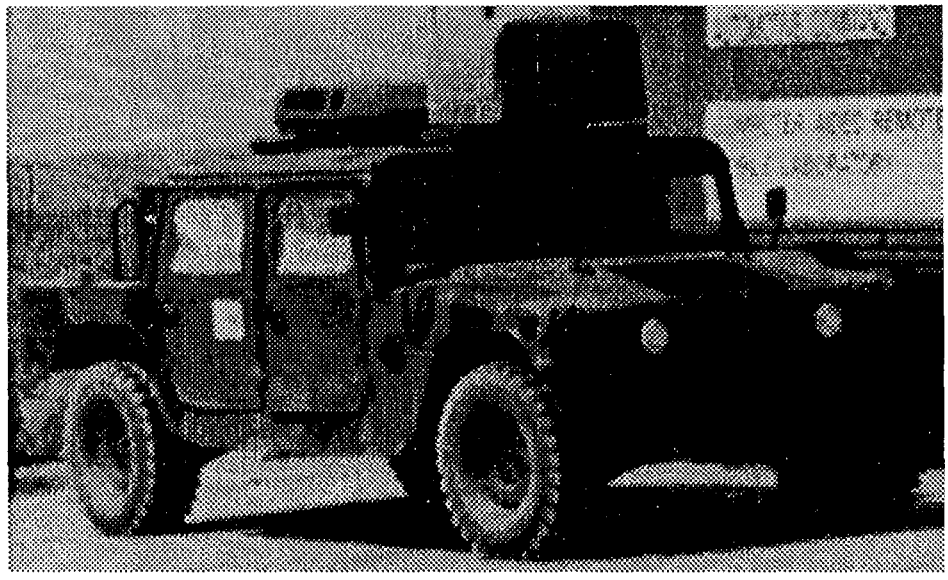


Figure 7: The JPL HMMWV testbed vehicle

and mounds of dirt (Sept '92) [1]. In this demonstration, the vehicle stopped at each frame.

- On the JPL HMMWV (figure 7), running at 64 x 60 resolution in ≈ 1.5 seconds/frame, for autonomous obstacle detection and halting during the U.S. Army "Demo I" exercise at Aberdeen Proving Ground (April'92). The HMMWV drove continuously at ≈ 1 to 3 m/see (≈ 5 to 10 kph) while detecting half-meter obstacles in time to stop.

- On a Martin Marietta HMMWV, integrated with the "Ganesha" and "Ranger" obstacle mapping and path planning systems developed at CMU [9, 10], running at 128 x 120 resolution in under one second per frame, for autonomous navigation on flat terrain with moderately tall grass and positive obstacles ≥ 1.0 m high ("Demo B", June'94). The vehicle velocity in this case was 2 m/see (7.2 kph).

We expect that the system will operate on a Martin Marietta HMMWV at "Demo C" in July' 95, running a 256 x 45 window of attention in about 0.5 seconds per frame. Other parameters of the demo remain to be determined. In preparation for this demo, the most recent versions of the stereo vision system and the Ranger path planning system were integrated on the JPL HMMWV and field tested in December'94. Figure 8 shows results from a test run that travelled about 200 m down the road shown in figure 2. The left side of the figure shows 256 x 45 windows of attention from eight positions along the run; the right side of the figure shows the corresponding range images. A composite elevation map of the entire run is shown in the middle; figure 9 shows a wireframe rendering of the composite map. The vehicle was instructed to follow its current heading whenever it could, but to swerve to a new heading as necessary to avoid obstacles. Two significant swerves are seen in this run. The first was to avoid a tree on the right side of the road; the second was to avoid a bush on the left side of the road, just beyond the tree. This was a very successful run, and typical of results obtained in this kind of terrain. Due to long computing latencies experienced with the existing computing system (figure 1), the vehicle speed was only 0.75 m/see. The computing system will soon be upgraded by putting a Spare-5CE CPU board in the same card cage as the

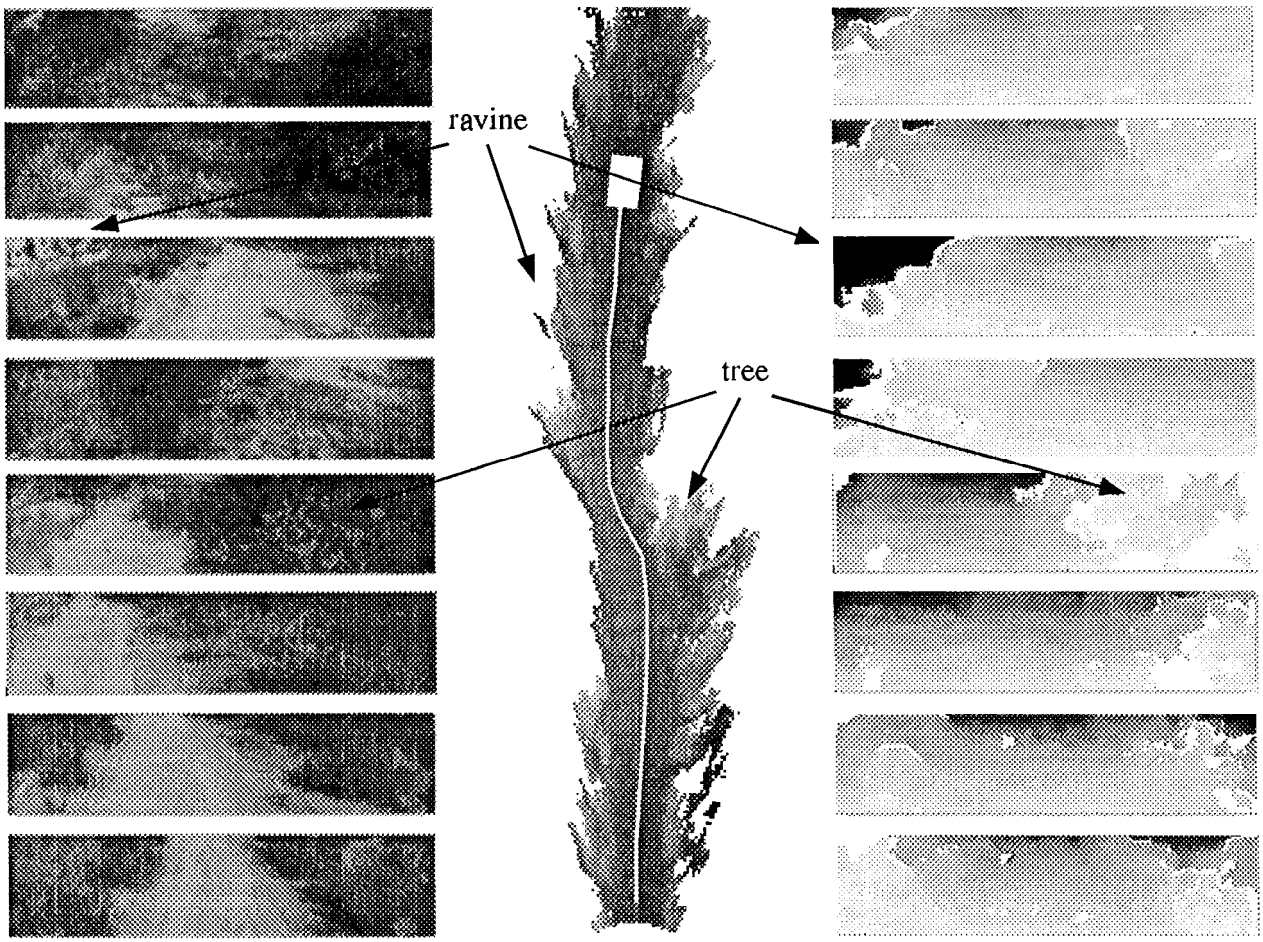


Figure 8: Road run with Ranger, covering about 200 meters. Shown are windows of attention from representative intensity images (left side), corresponding range images (right side), and a composite elevation map from the whole run (center). The white curve and rectangle on the elevation map represent the vehicle path and the vehicle itself at the end of the run.

68040. This will replace the Sparcbook, greatly reduce 1/0 delays, and provide at least a factor of three increase in computing speed.

5 Summary and Discussion

In this paper, we reviewed the current state of development of JPL's real-time stereo vision system, described ongoing development of new capabilities, and summarized significant field demonstrations that have used this system to perform autonomous navigation. The system currently produces range data from 256 x 45 windows of attention in approximately 0.6 seconds/frame. Ongoing development will extend the system to process imagery at512 x 480 resolution using full frame, non-interlaced cameras and a variety of techniques to mitigate the computational burden. Recent experiments with FLIR cameras have shown good results with imagery processed at 128 x 120 resolution; the results shown here were with thermal imagery collected by day, but qualitatively

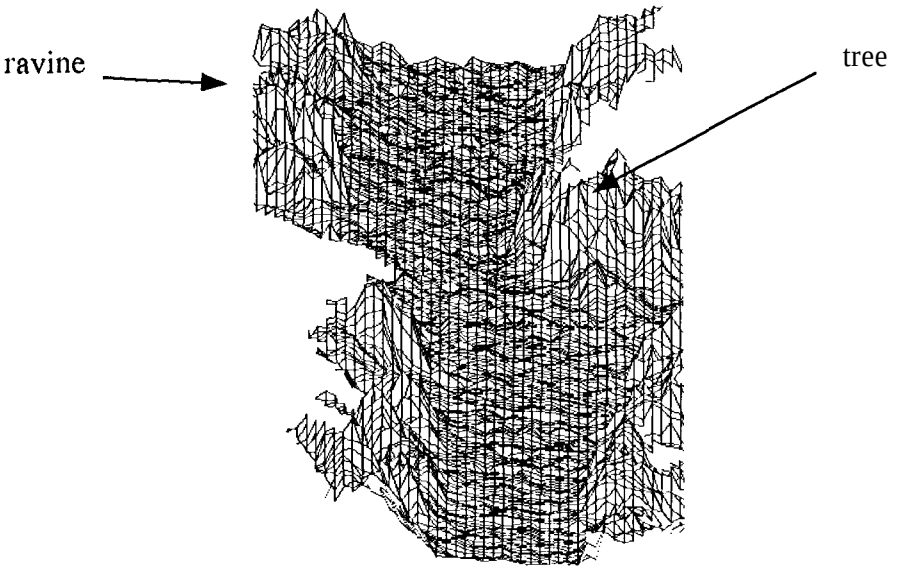


Figure 9: Wireframe rendering of the composite elevation map in figure 8.

similar results have been achieved at night. Field demonstrations have driven HMMWV's at speeds on the order of 5 to 10 kph over relatively flat, cross-country terrain covered with calf-high grass and sprinkled with positive obstacles 0.5 m high and larger.

This work has been undertaken with a pragmatic view toward demonstrating reliable performance with simple, fast algorithms. This strategy has paid off; it has shown that stereo vision can provide usable range data for UGV'S in a timely, relatively cost-effective manner and that stereo is a promising alternative to emissive range sensors, including laser scanners. For example, optimized, software-only versions of the algorithms described here are now planned to be used on prototype microrovers for Mars exploration; until recently, stereo was still considered too compu nationally expensive for such vehicles. Future work will improve the quality of the range data, integrate terrain classification capabilities with range imaging, and use new imager and computing technologies to reduce the mass, volume, and power consumption of the vision system.

6 Acknowledgements

Thanks to Bob Belles and Harlyn Baker of SRI International for bringing the FLIR cameras to JPL, as well as assisting in mounting them on the HMMWV, interfacing them to the vision system, calibrating them, and evaluating the results. Thanks to Terry Boult for helping to collect the multispectral imagery.

References

[1] L. II. Matthies. Stereo vision for planetary rovers: stochastic modeling to near real-time implementation. *International Journal of Computer Vision*, 8(1):71- 91, July 1992.